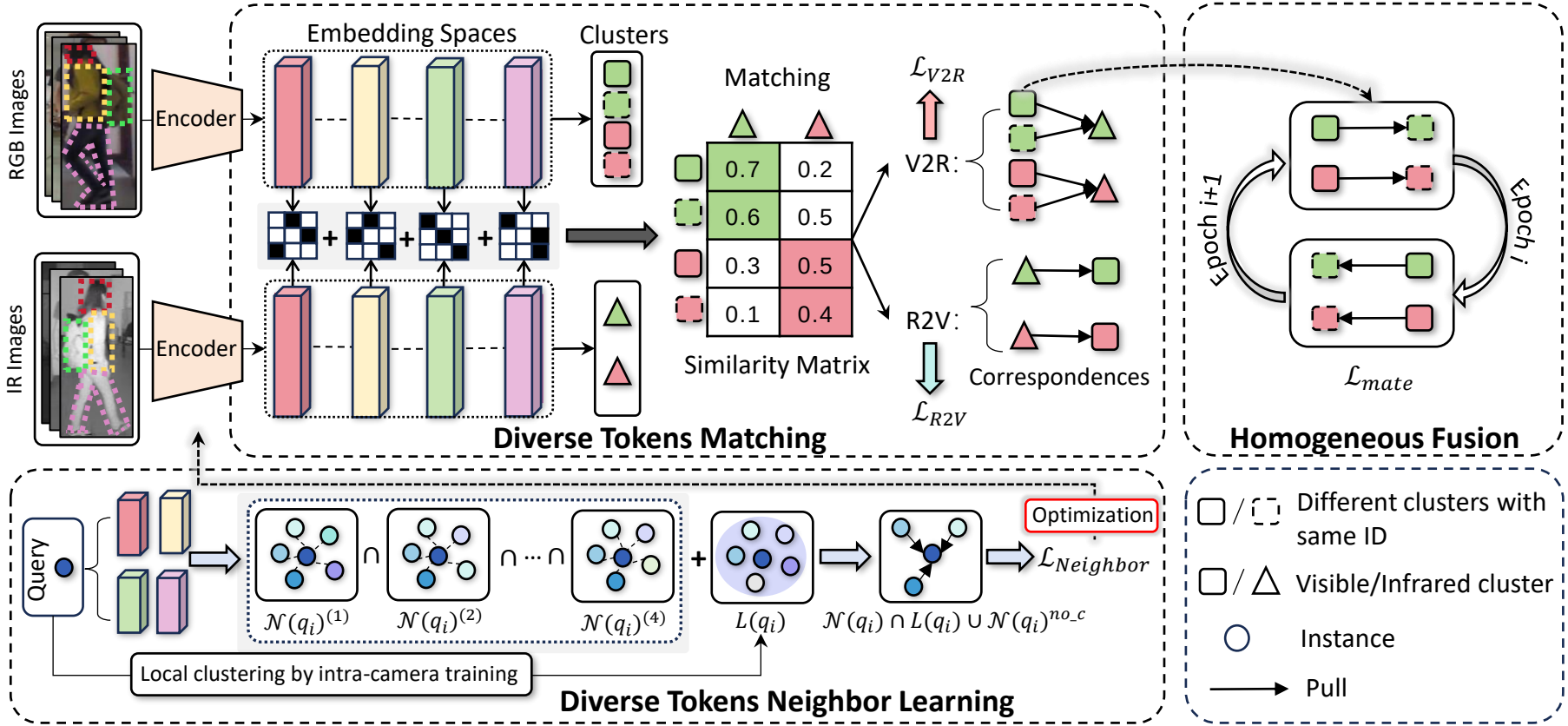
RGB Images
IR Images
Encoder
Encoder
Embedding Spaces
Clusters
Matching
0.7
0.2
0.6
0.5
0.3
0.5
0.1
0.4
Similarity Matrix
$\mathcal{L}_{V2R}$
V2R:
R2V:
Correspondences
$\mathcal{L}_{R2V}$
Diverse Tokens Matching
Epoch i+1
Epoch i
$\mathcal{L}_{mate}$
Homogeneous Fusion
Query
$\mathcal{N}(q_i)^{(1)}$
$\mathcal{N}(q_i)^{(2)}$
$\mathcal{N}(q_i)^{(4)}$
$L(q_i)$
$\mathcal{N}(q_i) \cap L(q_i) \cup \mathcal{N}(q_i)^{no\_c}$
Optimization
$\mathcal{L}_{Neighbor}$
Local clustering by intra-camera training
Diverse Tokens Neighbor Learning
Different clusters with same ID
Visible/Infrared cluster
Instance
Pull